Response to reviewer 2

We begin by thanking the reviewer for their very helpful comments.

However, I am concerns that this manuscript is insufficient to be useful due to lack of substantial materials and logical reasoning in current version.

We have added explanations to substantiate some of the results. For example, we now discuss a possible cause for why MISR does not capture the hotspot in climatological AOD as well as the other instruments. We have provided a reason why median AOD and $PM_{2.5}$ mass densities are preferable for the spatial correlation analysis in the revised manuscript (as opposed to mean values used in the original manuscript). We feel the discrepancy in long-term trends between the satellite sensors is not strong, but now suggest that the MODIS calibration degradation could account for the general negative trend in AOD from this sensor. Further details are provided below on each issue. This is simply a summary of our response.

First of all, I have read the comments from Andrew Sayer, who is an expert on aerosol retrievals from satellite-based remote sensing, especially in MODIS AOD retrievals. His comments are very useful to improve the understanding of the MODIS AOD retrievals and improve current studies.

Dr. Sayer's comments have helped to improve the revised manuscript. The reviewer can refer to our response to Dr. Sayer's comments to see the resulting changes to the manuscript.

My major concerns about this manuscript are the lack of in-depth analysis and lack the necessary explanations. For example, the finding of the ability to capture spatial variability with MISR is generally much worse than the other instruments over AOSR region is very interesting and useful to know the limitation of MISR measurements, however the possible reasons for this will be more important to see the spatial limitation of MISR.

The MISR spatial limitation, evident in Fig. 1, is probably due to its spatial sampling being tied to its temporal sampling. We found locations within the AOSR where MISR was measuring almost exclusively in October. Thus, the seasonal cycle in AOD is aliasing into the AOD spatial distribution.
The spatial correlation coefficient is based on 10 sites. Because of the small number of sites, the correlation is quite sensitive to a bias in AOD or $PM_{2.5}$ at any station. Wapasu has significantly higher mean $PM_{2.5}$ mass density for MISR coincidences than any other site (10.2 $\mu g/m^3$ while the next highest site average is 8.1 $\mu g/m^3$). MISR overpasses of Wapasu span only two years (2014-2015) and these years were affected by anomalously high forest fire activity in western Canada. The median reduces the sensitivity to these outliers as compared to the mean. In the revised manuscript, Table 3 now contains the correlation of the median of coincident $PM_{2.5}$ and satellite AOD data. This table is inserted below. The revised Table 3 shows the spatial correlation coefficient (R) of MISR AOD with $PM_{2.5}$ is not much worse than the spatial R of MODIS/Aqua DT and $PM_{2.5}$.

AOD product	R	N
POLDER/PARASOL 865 nm	0.64	8
AATSR 550 nm	0.73	9
MISR 558 nm	0.20	10
MODIS/Aqua DT 550 nm	0.23	10
MODIS/Aqua DB 550 nm	0.57	10

At p3L34, we now modify the description of the spatial correlation analysis as follows:

In order to assess the ability of the satellite data to capture the spatial variability in this region, the hourly in-situ surface-level $PM_{2.5}$ from the 10 NAPS (National Air Pollution Surveillance) stations (Table 2) are used. Demerjian (2000) provided a review of the NAPS network, but since 2011, this network has undergone a gradual shift in the continuous monitoring of $PM_{2.5}$ mass density from tapered element oscillating microbalances (TEOMs) to the SHARP (Synchronized Hybrid Ambient Real-time Particulate) monitoring system. The latter is a hybrid system, consisting of a nephelometer and a beta attenuation monitor (Hsu et al., 2016). The spatial correlation between median satellite AODs and NAPS $PM_{2.5}$ mass densities is determined using coincident data.

At p5L4, we now update the text with the following:

The AOD hotspot in the AOSR seen by POLDER is less obvious with MISR (Fig. 1). This is consistent with the relatively poorer ability of MISR to capture spatial variability based on spatial correlations of median AOD and $PM_{2.5}$ mass density over the ~10 available sites (Table 3). While the spatial correlation analysis relies on temporally coincident data, the less obvious AOD hotspot for MISR in Fig. 1 is also partly due to the spatio temporal sampling by this instrument. Some locations are only sampled during a short period of the year, and thus the seasonal cycle of AOD is aliased into the MISR spatial distribution.

In section 3.1, the authors have indicated that all of the satellite retrievals can capture the inter-annual variability of the annual mean AOD observed by AERONET, but the trends estimated based on the each satellite retrievals showed lots of differences, some of positive and some of negative. Thus, what are the main reasons to explain this discrepancy?

We agree that there is a discrepancy between the trends estimated by the different satellite AOD products, but it is not strong. The satellite data records all span approximately one decade. A period of a decade is rather short for determining a trend, considering the natural interannual variability in AOD and possible instrumental drifts (e.g. Levy et al., 2015). Focussing on the Muskeg River mining region where there appears to be a significant positive AOD trend according to MODIS/Aqua DB and POLDER/PARASOL, the linear trend is not different from zero for both AATSR and MISR (p6L28-29). Also, MODIS/Aqua DT has a slightly negative trend, but it is also not different from a null trend, so given that none of AOD products show a strong decreasing trend in this Muskeg River mining region, there is no strong discrepancy in the AOD trends. The insignificant negative AOD trend for MODIS/Aqua DT remains now that we have switched to 550 nm.

We now add at p8L1:

The calibration of the MODIS reflective solar bands is achieved by calibration with the solar diffuser. Some negative drift in AOD (Levy et al., 2015) is expected for MODIS Aqua similar to its Terra counterpart (see Sect. 2) as the designs of the solar diffuser and its stability monitor are nearly identical in the two MODIS sensors (Wu et al., 2013).

The authors reported a major issue of satellite AOD retrievals over this region, which is the lack of successful retrieval samples, especially of the MODIS retrievals which has low confidence. It is good information. However, the reasons for the large part of retrievals has low confidence are not well explained.

The reasons for the low confidence of MODIS AODs were explained in the ACPD version of the manuscript (p5L24-26 for Deep Blue and p5L12-19 for Dark Target). An additional reason for MODIS Dark Target has been added to the revised manuscript: coastal areas (see comment by Dr. Sayer and response).

Furthermore, the comparison of coincident AODs observed by satellite-based and AERONET shows large bias (more than 20%) between them, but necessary explanations are not provided.

MISR is the only satellite-based aerosol sensors with a consistent bias of >20% in this region. Explanations were included in the ACPD version (p9L5-11), although one literature reference has been updated in these sentences.

I found that the correlation between monthly mean of the satellite retrieved AOD and AERONET AOD are analyzed, but I'd suggest to use the individual samples from AERONET to evaluate the satellite AOD retrievals and discuss the bias of each satellite product.

This is already done in Table 4. The second to fourth columns in Table 4, namely 'r_s', 'slope', and 'offset', are all based on individual coincidences. Although it can be inferred from the ACPD version of the manuscript that the quantities in these columns are based on a regression using individual coincidences (e.g. p1L12 and p9L3-4), we will be more explicit in Sect. 2. At p3L30, we now write

"Since individual AERONET and satellite AODs are not normally distributed, we use linear least-squares weighted by Huber's function to determine the slope and offset since this is a robust method that does not completely disregard highly deviating points (Bergström and Edlund, 2014). (…) Similarly, due the non-normal distribution of the individual AOD data, Spearman's rank correlation coefficient (r_s) is chosen to study the site-specific AOD correlation based on individual AERONET-satellite coincidences."

In the conclusion (p8L17), we now repeat that correlation was determined using individual AERONET observations:

"However, the MODIS dark target product is the best at capturing temporal variability in terms of the correlation with individual AERONET AODs at Fort McMurray…"

It is not clear to describe how to derive the PM2.5 mass density from satellite AODs. I noticed that the constant ratio of PM2.5 to AOD is used to convert the AOD trends from satellite instruments to PM2.5 trends. However, this is not accurate. The relationship between surface PM2.5 and AOD is not always linear. It is affected by multiple factors, such as the relative humidity since the AOD can be enhanced by aerosol swelling effects but the PM2.5 does not. Meanwhile, the correlation between AOD and surface level PM2.5 significantly depends on the aerosol vertical distribution and aerosol particle size distribution. Thus, the uncertainties in those analysis and the influences on the results should be discussed.

The existing manuscript was not clear about the timescale when the word "constant" was used. What was meant is that the $PM_{2.5}$/AOD ratio is assumed to be constant from year to year (based on annually averaged ratios). This ratio can even change from year to year if there were an increasing trend in surface-level aerosol emissions. In the revised manuscript, we have devised a better way to compare trends: the POLDER/PARASOL and MODIS Deep Blue AOD offsets, determined from the AERONET validation at Fort McMurray, are corrected and then relative trends are used for $PM_{2.5}$ and satellite AOD. Thus, the $PM_{2.5}$/AOD ratio is not used in the revised manuscript. The Fort McMurray AERONET site is used for bias correction since it has temporal overlap with both sensors and has a longer record than the Fort McKay site. There is qualitative agreement on the magnitude of the offset at both sites for MODIS DB.

This is now described at p4L12:

For temporal trends, a simple linear regression is performed on relative anomalies derived from bias-corrected annual average and median AODs. The bias correction involves subtracting the AOD offset obtained through the validation with coincident Fort McMurray AERONET data.

P6, Line 28: Is this trend statistical significant?

Yes, the MODIS/Aqua DB and POLDER/PARASOL trends are both statistically significant. We will add "statistically" to the sentence as follows:

In fact, two satellite data products, namely POLDER/PARASOL and MODIS/Aqua DB, exhibit a statistically significant positive trend in this mining area.